

Fri Mar 14 14:00:07 2003

us-09-698-781-3.oli.rst

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM Protein - nucleic search, using frame_plus.p2n model

Run on: March 14, 2003, 05:42:09 ; Search time 1720 Seconds

(without alignments)
2429.323 Million cell updates/sec

Title: US-09-698-781-3

Perfect score: 258

Sequence: 1 MKOILHFALETTAMTLPVL.....KHQVLRSCASCSNCNSIY 258

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Xgapop 60.0 , Xgapext 60.0
Xgapop 6.0 , Xgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 341630

Minimum DB seq length: 20
Maximum DB seq length: 99

Post-processing: Listing first 100 summaries

Command line parameters:

-MODEL-frame+.p2n.model -DEV-xih
-O/cqnt2.1/USPTO.spool/US09698781/runat_07032003_083810.6968/app.query.fasta_1.455
-DB-EST -OPTM-fastap -SUFFIX-oli.rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS-bits -STRAT=1 -END=1 -MATRIX-oligo -TRANS-human40.cdi -LIST=100
-DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=30 -MODE=LOCAL -OUTFMT=pro
-USPR=US09698781.ecgn_1.1906.efunc_07032003_083810.6968 -NCPU=6 -ICPU=3
-NO_ALIGN -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WANT_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	7.0	55	9	AA661880 nt95106.s
2	7	2.7	32	17	AZ661946 2M0230002
3	7	2.7	81	14	BT061111 EBR04.SO
4	7	2.7	92	14	T89904 yella09.rl
5	7	2.7	92	17	BH638251 1008021D0
6	7	2.7	99	9	AA143118 zof98a11.r
7	6	2.3	31	9	AI142775 qra6402.s
8	6	2.3	32	17	TA18D05P
9	6	2.3	35	17	AZ838204 2M0133K17
10	6	2.3	40	9	AA726265 vub9c01.r
11	6	2.3	41	14	H94066 yv15h02.s1
12	6	2.3	41	17	AZ813416 2M0080M10
13	6	2.3	46	17	AZ412916 1M0166J11
14	6	2.3	48	14	C21082 HUMS000259
15	6	2.3	49	9	AI197659 ued6e09.r
16	6	2.3	50	9	AJ281339 4A3A-P2A7
17	6	2.3	50	9	AU102979 AU102979
18	6	2.3	51	9	AA616601 vrb6h08.r
19	6	2.3	51	17	BH218157 100607D0
20	6	2.3	52	17	B00049 CSRL-100h6-
21	6	2.3	58	17	AL733018 Danilo rer
22	6	2.3	58	17	AL758979 Arabidops
23	6	2.3	58	17	AZ316082 1M0033P01
24	6	2.3	60	17	AZ313488 1M0029E08
25	6	2.3	62	12	BG095925 nag53c09.
26	6	2.3	62	12	BF037661 601462102
27	6	2.3	63	17	AZ500262 1M0338H04
28	6	2.3	63	9	AA761487 n22th09.s
29	6	2.3	63	17	AF219092 AF219092
30	6	2.3	64	13	B1386670 EST-CD34N
31	6	2.3	64	17	AZ479811 1M0300116
32	6	2.3	65	17	AL766062 Arabidops
33	6	2.3	66	17	AZ331997 1M0060E09
34	6	2.3	67	17	AZ617135 1M00448H13
35	6	2.3	68	17	BH804876 1008105F0
36	6	2.3	69	17	BH638602 1008023D1
37	6	2.3	72	13	BM424830 IPSPO0073
38	6	2.3	73	17	BH853910 R89821
39	6	2.3	73	17	BH860880 Gm-Stub00
40	6	2.3	73	17	BH853910 SALK_0784
41	6	2.3	73	17	BH862853 SALK_0907
42	6	2.3	74	9	AA854384 aj76b03.s
43	6	2.3	74	14	C21414 HUMS000947
44	6	2.3	75	17	AZ506280 2M0148E05
45	6	2.3	75	17	AZ847788 2M0148E05
46	6	2.3	77	17	AL762744 Arabidops
47	6	2.3	76	14	AI540827 tp68a12.x
48	6	2.3	76	14	B0816306 1003056G0
49	6	2.3	77	13	AZ480476 1M0302102
50	6	2.3	78	17	B1517667 603042278
51	6	2.3	78	17	BH847206 SALK_0445
52	6	2.3	79	9	A1789866 u65e60.r
53	6	2.3	79	14	B0094073 040802_32
54	6	2.3	79	14	B0094073 040802_50
55	6	2.3	79	14	B0094091 040802_51
56	6	2.3	80	9	AA889157 ak25d06.s
57	6	2.3	80	17	CNS01X08
58	6	2.3	82	9	AA283715
59	6	2.3	82	10	AA914126 AV914126
60	6	2.3	83	14	B0789779 haeg002AB
61	6	2.3	83	17	BH221159 100609G00
62	6	2.3	85	9	AJ283196 4A3A-PG3
63	6	2.3	85	13	B1758902 603042352
64	6	2.3	85	17	BH802812 1008102E0

```

c 65 6 2.3 86 9 AA552526
c 66 6 2.3 86 10 AM262304
c 67 6 2.3 86 14 BQ625486
c 68 6 2.3 86 17 A2466721
c 69 6 2.3 86 17 A2759836
c 70 6 2.3 88 9 AA871948
c 71 6 2.3 88 9 F30846
c 72 6 2.3 89 17 AF219091
c 73 6 2.3 89 17 CDS02RCD
c 74 6 2.3 90 9 AA971814
c 75 6 2.3 90 9 AF027915
c 76 6 2.3 90 9 A1470812
c 77 6 2.3 90 17 A2921474
c 78 6 2.3 91 9 AA922763
c 79 6 2.3 91 9 A1968724
c 80 6 2.3 91 13 B1557272
c 81 6 2.3 91 17 A2374702
c 82 6 2.3 93 14 BQ376513
c 83 6 2.3 93 17 AL764388
c 84 6 2.3 94 9 A281343
c 85 6 2.3 94 9 A283110
c 86 6 2.3 94 17 A2941059
c 87 6 2.3 94 17 BH612035
c 88 6 2.3 95 9 A1664297
c 89 6 2.3 95 17 A2603988
c 90 6 2.3 96 9 AA012296
c 91 6 2.3 96 10 AM059816
c 92 6 2.3 96 12 BG111920
c 93 6 2.3 96 17 A2465383
c 94 6 2.3 96 17 BH224140
c 95 6 2.3 97 9 A1583847
c 96 6 2.3 98 9 AB036296
c 97 6 2.3 98 9 A0177039
c 98 6 2.3 98 14 BQ566189
c 99 6 2.3 98 17 AL766046
c 100 6 2.3 98 17 CDS0308K

```

ALIGNMENTS

```

RESULT 1
AA661880 55 bp mRNA linear EST 16-DEC-1997
LOCUS nt95f06.s1 NCI-CCAP_P12 Homo sapiens cDNA clone IMAGE:1206275
DEFINITION similar to TR:E221223 E221223 CYSTEINE-RICH SECRETORY PROTEIN-3
PRECURSOR. (1) ; mRNA sequence.
ACCESSION AA661880
VERSION AA661880.1 GI:2615971
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens

```

```

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
1 (bases 1 to 55)
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM, at:
www-bio.llnl.gov/dbtrp/image/image.html

```

Trace considered overall poor quality
Insert length: 364 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amerisham
High quality sequence stop: 1.

```

FEATURES
Source Location/Qualifiers
1..55
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1206275"
/clone_id="NCI-CCAP_P12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: PAMP10: mRNA made by metastatic prostate
lesion of the bone. cDNA made by oligo-dT priming.
Non-directionally cloned. Size selected on agarose gel,
average insert size 600 bp. Library made by D. Kitzman,
NIH."
BASE COUNT 15 a 11 c 13 g 16 t
ORIGIN
Alignment Scores:
Pred. No.: 8.14e-09 Length: 55
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.98% Indels: 0
DB: 9 Gaps: 0
US-09-698-781-3 (1-258) x AA661880 (1-55)
QY 188 G1YASnTTPAlaAsnArgLeuTYrValProTYrGInGInGInAlaProCysAla 205
Db 2 GGTAAATTGGGCTAATGACATATATGTCCTTATGAAACAGACACCTTGTC 55
RESULT 2
AA2961946 32 bp DNA linear GSS 27-APR-2001
LOCUS clone UUCG2M0230002 R. DNA sequence.
DEFINITION
ACCESSION AA2961946
VERSION AA2961946.1 GI:13833173
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
REFERENCE Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 32)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0230 row: 0 column: 02
Seq primer: CACACAGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 32.
FEATURES
Source Location/Qualifiers
1..32
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone_image="UUCG2M0230002"
/clone_id="Mouse 10kb plasmid UUCG2M library"
/sex="female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

```

/note="vector: pMD22uv. Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD22 (914/732114/9D/RT1902.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 3 c 13 g 13 t

ORIGIN

Alignment Scores:

Pred. No.:	417	Length:	32
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.71%	Indels:	0
DB:	17	Gaps:	0

US-09-698-781-3 (1-258) x A2961946 (1-32)

Oy 22 PheleuValaMaglyleu 28

Db 3 TTTCTTACTGCTTATTG 23

RESULT 3

BO761111

LOCUS BO761111 81 bp mRNA linear EST 26-JUL-2002

DEFINITION EBR04_S0004_B23_R root, 3 week, salt-stressed, cv Optic, EBR04

Hordeum vulgare cDNA clone EBR04_S0004_B23 5', mRNA sequence.

ACCESSION BO761111.1 GI:21969583

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: estescr1.scri.ac.uk
Location/Qualifiers

FEATURES

source

1. 81
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBR04_S0004_B23"
/clone_1lb="root, 3 week, salt-stressed, cv Optic, EBR04"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/note="vector: pSPORT1, site 1: Sal I, site 2: Not I;
Non-normalised library, directionally cloned into pSPORT1."

Derived from roots of 3 week old salt stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project.

BASE COUNT 18 a 25 c 26 g 12 t

ORIGIN

Alignment Scores:

Pred. No.:	1.16e+03	Length:	81
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.71%	Indels:	0
DB:	14	Gaps:	0

US-09-698-781-3 (1-258) x BO761111 (1-81)

Oy 151 HistfThrglnvalValTrp 157

Db 59 CACTACACACAGCTGCTGCG 79

RESULT 4

LOCUS

DEFINITION T89904 92 bp mRNA linear EST 20-MAR-1995

IMAGE:117400 5', mRNA sequence.

ACCESSION T89904

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Chissos, S., Dietrich, N., Dubuque, T., Favello, R., Gish, M., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Marquis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Riklin, L., Rohlfing, T., Schellendberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Wolfdamm, P., Waterston, R., Wilson, R., and Marra, M. Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estescr1.wustl.edu
Insert Size: 1079
High quality sequence stops: 62 Source: IMAGE Consortium, LINT This clone is available royalty-free through LINT; contact the IMAGE Consortium (infocenter@imgl.gov) for further information.
Insert Length: 1079 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 62.
Location/Qualifiers

FEATURES

source

1. 92
/organism="Homo sapiens"
/db_xref="GDB:485689"
/db_xref="taxon:9606"
/clone="IMAGE:117400"
/clone_1lb="Stratagene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: Lung; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT, normal lung. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GATTGGGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
BASE COUNT 34 a 21 c 24 g 13 t

Alignment Scores:

Pred. No.:	1.34e+03	Length:	92
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.71%	Indels:	0
DB:	14	Gaps:	0

US-09-698-781-3 (1-258) x T89904 (1-92)

OY 25 AlaglyLeuProSerPhe 31
 |||||
 Db 86 GCGGCTTCCTCTCTTC 66

RESULT 5
 BH638251

LOCUS 1008021D04.1EL_Y1 1008 - Rescemu Grid 1 Zea mays genomic, DNA

DEFINITION 92 bp DNA linear GSS 14-FEB-2002

ACCESSION BH638251

VERSION BH638251.1 GI:18662118

KEYWORDS GSS.

SOURCE Zea mays.

ORGANISM Zea mays.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

AUTHORS 1 (bases 1 to 92)

TITLE Walbot V

JOURNAL Maize genomic sequences found using engineered Rescemu transposon unpublished (2001)

COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 723 8221
 Email: walbot@stanford.edu
 Very probable ligation site of ends cut by single endonuclease.
 Reverse complemented post-ligation sequence from source sequence.
 Plate: 1008021 row: 9
 Class: transposon tagged.
 Location/Qualifiers

FEATURES

source

1..92
 /organism="Zea mays"
 /cultivar="mixed background W23/A18/B73"
 /db_xref="taxon:4577"
 /clone_11b="1008 - Rescemu Grid 1"
 /tissue_type="leaf"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: leaf; Vector: Rescemu (engineered from plasmid backbones); Site_1: BamHI; Site_2: BglII; Rescemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescemu, go to the web site www.zmmb.iastate.edu and follow the links for 'Rescemu'. Grid 1 was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 20 a 21 c 19 g 32 t

ORIGIN

Alignment Scores:

Pred. No.:	1.34e+03	Length:	92
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.71%	Indels:	0

DB: 17 Caps: 0

US-09-698-781-3 (1-258) x BH638251 (1-92)

OY 16 LeupheProValLeuPhe 22
 |||||
 Db 20 CTATTCAGTCTACTATT 40

RESULT 6
 AA143118

LOCUS 99 bp mRNA linear EST 08-NOV-1997

DEFINITION z069ell.r1 StrataGene pancreas (#937208) Homo sapiens cDNA clone IMAGE:592172.5, similar to gb:X57138_rnal HISTONE H2B.2 (HUMAN);, mRNA sequence.

ACCESSION AA143118

VERSION AA143118.1 GI:1712497

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 99)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theisling,B., White,X., Wyllie,T., Waterston,R. and Wilson,R.

TITLE WashU-NCI human EST Project

JOURNAL Unpublished (1997)

COMMENT Contact: Wilson R.
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Trace considered overall poor quality
 Insert Length: 622 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amsterdam
 High quality sequence stop: 1.
 Location/Qualifiers

FEATURES

source

1..99
 /organism="Homo sapiens"
 /db_xref="GDB:4623558"
 /db_xref="taxon:9606"
 /clone_11b="IMAGE:592172"
 /clone_11b="Stratagene pancreas (#937208)"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: pancreas; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAC 3' -3' adaptor sequence: 5' CTCGACGTTTTTTTTTTT 3'."
 CTCGACGTTTTTTTTTTT 3'."

BASE COUNT 37 a 21 c 27 g 10 t 4 others

ORIGIN

Alignment Scores:

Pred. No.:	1.45e+03	Length:	99
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.71%	Indels:	0
DB:	9	Gaps:	0

US-09-698-781-3 (1-258) x AA143118 (1-99)

OY 202 AlaprocysAlasercysPro 208
 |||||
 Db 66 GCTCGCTGCGCTCTCTCCG 46

RESULT 7
 AA142775/c

LOCUS A1142775 31 bp mRNA EST 23-OCT-1998
 DEFINITION q926a02.s1 Soares_NHMPU_S1 Homo sapiens CDNA clone IMAGE:1687850
 ; similar to TR:Q05519 Q05519 ARGININE-RICH 54 KD NUCLEAR PROTEIN.
 ; mRNA sequence.
 ACCESSION A1142775
 VERSION A1142775.1 GI:3659134.
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 31)
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LINT: contact the
 IMAGE Consortium (infoimage.lint.gov) for further information.
 Trace considered overall poor quality
 Insert Length: 1957 Std Error: 0.00
 Seq primer: -40m13 fwd. Ex from Amer sham
 High quality sequence stop: 1.
 FEATURES
 source
 1. 31
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1687850"
 /clone_1b="Soares_NHMPU.S1"
 /clone_type="Pooled human melanocyte, fetal heart, and
 /lab_host="DH108"
 /note="Organ: mixed (see below); Vector: pTZ19-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not 1;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NBH, pregnant uterus
 NBHP, and fetal heart NBH19W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of 1 M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."
 BASE COUNT 7 a 12 c 8 g 4 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4e+03 Length: 31
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 9 Gaps: 0
 US-09-698-781-3 (1-258) x A1142775 (1-31)
 OY 24 ValA1aglyLeuLeuPro 29
 LOCUS |||||||
 Db 25 GTGGCAGGCTCTCTGCT 8
 RESULT 8
 TA318D05P 32 bp DNA linear GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 318D05, forward sequence.
 DEFINITION genomic survey sequence.
 ACCESSION AL492652
 VERSION AL492652.1 GI:11867479
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.

REFERENCE 1 (bases 1 to 32)
 AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
 Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
 Melville, S.E., Rajandream, M.A. and Barrell, B.G.
 TITLE Direct Submission
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridgeshire CB10 1SA. E-mail: barrell@sanger.ac.uk and
nhs@sanger.ac.uk
 COMMENT Constructed at the Institute for Genomic Research (TIGR),
 Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
 to give a tight size distribution (4 kb). The v + i method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999).
 Email: neilsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available
 at <http://www.sanger.ac.uk/projects/T-brucei/>.
 FEATURES
 source
 1. 32
 Location/Qualifiers
 /organism="Trypanosoma brucei"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="318D05"
 BASE COUNT 8 a 6 c 5 g 13 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.15e+03 Length: 32
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 17 Gaps: 0
 US-09-698-781-3 (1-258) x TA318D05P (1-32)
 OY 228 AsnCysLySerLeuLys 233
 LOCUS |||||||
 Db 23 AACTGTAAGGCTTAAG 6
 RESULT 9
 A2838204/c 35 bp DNA linear GSS 20-FEB-2001
 LOCUS 2M0133K17R Mouse 10kb plasmid UGCC1M library Mus musculus genomic
 DEFINITION clone UGCC2M0133K17 R. DNA sequence.
 ACCESSION A2838204
 VERSION A2838204.1 GI:13008112
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathu; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 35)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Bacon, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Federsen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00

Plate: 0133 row: K column: 17
 Seq primer: CACACAGCAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 35.
 Location/Qualifiers

FEATURES

source

1..35
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U06C2M033K17"
 /clone_1lb="Mouse 10kb plasmid U06C1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMDA2ny. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMDA2 (g11473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 11 a 0 c 24 g 0 t

Alignment Scores:

Pred. No.: 4.58e+03 Length: 35
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x A2838204 (1-35)

OY 27 LeuLeuProSerPhePro 32

DB 20 CTCCTCCCTCCCTCCCT 3

RESULT 10

AA726265

LOCUS

DEFINITION

AA726265 40 bp mRNA linear EST 02-JAN-1998
 v889c01.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:1209312 5' similar to TR:008634 008634 HIGH-SULFUR KERATIN
 PROTEIN ; mRNA sequence.

ACCESSION

AA726265

VERSION

AA726265.1 GI:2743972

KEYWORDS

EST.

ORGANISM

house mouse.

REFERENCE

Mus musculus

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 40)

TITLE

Marr M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,

JOURNAL

Genes S., Kucaba T., Lacey M., Le M., Martin J., Morris M.,

COMMENT

Schellenberg R., Stepien M., Tan F., Underwood R., Moore B.,

Washington University School of Medicine

FEATURES

source

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mousestewatson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:645656
 putative full length read
 vector to vector length is 592
 Seq primer: -28m13 rev1 EF from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers

1..40
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone_1lb="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
 ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
 dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."
 BASE COUNT 3 a 18 c 11 g 8 t

Alignment Scores:

Pred. No.: 5.31e+03 Length: 40
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 9 Gaps: 0

US-09-698-781-3 (1-258) x AA726265 (1-40)

OY 65 AlaValSerProProAla 70

DB 18 GCAGTGTCCACCGCGCC 35

RESULT 11

H94066

LOCUS

DEFINITION

H94066 41 bp mRNA linear EST 25-NOV-1996
 yv15h02.s1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone
 IMAGE:242833 3' similar to gb:Y00493.fna1 HEMOGLOBIN ALPHA CHAIN
 (HOMAN); mRNA sequence.

ACCESSION

H94066

VERSION

H94066.1 GI:1101362

KEYWORDS

EST.

ORGANISM

human.

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 41)

AUTHORS

Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman

TITLE

'M., Hultman M., Kucaba T., Le M., Lennon G., Marr M., Parsons J.,

JOURNAL

Ritkin L., Rohlfing T., Soares M., Tan F., Trevisan E., Waterston

COMMENT

The WashU-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence starts: 1
 High quality sequence stops: 1

Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Trace considered overall poor quality
 Insert Length: 1147 Std Error: 0.00
 Seq primer: m3 -40 forward.
 Location/Qualifiers

FEATURES

source

1. 41
 /organism="Homo sapiens"
 /db_xref="Gene:3791968"
 /db_xref="taxon:9606"
 /clone="IMAGE:242835"
 /clone_lib="Scores fetal liver spleen INTUS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
 with a modified polylinker. Site 1: Pac I; Site 2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5' ACTGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 9 a 9 c 10 g 11 t 2 others
 ORIGIN

Alignment Scores:

Pred. No.: 5.45e+03 Length: 41
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 14 Gaps: 0

US-09-698-781-3 (1-258) x H94066 (1-41)

OY 230 LysSerLeuLysLeuThr 235

Db 2 AAGACCTTGAACTTGACC 19

RESULT 12

A2813416

LOCUS

DEFINITION 2M0808M10R Mouse 10kb plasmid UGCC1M library Mus musculus genomic
 clone UGCC2M0808M10 R, DNA sequence.

ACCESSION

A2813416

VERSION

A2813416.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

41 bp DNA linear GSS 20-FEB-2001
 A2813416
 2M0808M10R Mouse 10kb plasmid UGCC1M library Mus musculus genomic
 clone UGCC2M0808M10 R, DNA sequence.
 A2813416
 A2813416.1 GI:12983324
 GSS.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 41)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel.: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0080 Row: M Column: 10
 Seq primer: CACACAGGAACGCTATGACC

Class: plasmid ends
 High quality sequence stop: 41.
 Location/Qualifiers

FEATURES

source

1. 41
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UCC2M0808M10"
 /clone_lib="Mouse 10kb plasmid UGCC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /note="Vector: pMD42ny; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g1147321149b1AR129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."
 BASE COUNT 7 a 13 c 8 g 13 t
 ORIGIN

Alignment Scores:

Pred. No.: 5.45e+03 Length: 41
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x A2813416 (1-41)

OY 231 SerLeuLysLeuThrLeu 236

Db 21 TCCCTCAGCTGACCTTG 38

RESULT 13

A2412916

LOCUS

DEFINITION 1M0186J11R Mouse 10kb plasmid UGCC1M library Mus musculus genomic
 clone UGCC1M0186J11 R, DNA sequence.

ACCESSION

A2412916

VERSION

A2412916.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

46 bp DNA linear GSS 03-OCT-2000
 A2412916
 1M0186J11R Mouse 10kb plasmid UGCC1M library Mus musculus genomic
 clone UGCC1M0186J11 R, DNA sequence.
 A2412916
 A2412916.1 GI:10536929
 GSS.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 46)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel.: 801 585 5606

Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0186 row: J column: 11
 Seq primer: CACACAGCAACACGATATACCC
 Class: plasmid ends
 High quality sequence stop: 46.
 Location/Qualifiers
 1..46

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U08C1M0186J11"
 /clone_1lb="Mouse 10kb plasmid U08C1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD29v, Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g114732114[9b]AF13902.1), a copy number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
 BASE COUNT 8 a 16 c 9 g 13 t
 ORIGIN

Alignment Scores:

Pred. NO.: 6.19e+03 Length: 46
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2,33% Indels: 0
 DB: 17 Gaps: 0

US-09-698-781-3 (1-258) x A2412916 (1-46)

OY 27 LeuLeuProSerPhePro 32

DB 27 CTGCTCCCTCTCCCT 44

RESULT 14

C21082/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

C21082 48 bp mRNA linear EST 23-OCT-1996
 HUMS0002596 Human adult (K.Okubo) Homo sapiens CDNA 3', mRNA
 sequence.
 C21082
 C21082.1 GI:1622192
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 48)
 Okubo, K.
 BodyMap: human gene expression database
 Unpublished (1995)
 Contact: Okubo, K.
 Institute for Molecular and Cellular Biol
 Osaka University
 1-3, Yamada-oka, Suita, Osaka Pref. 565, Japan
 Tel: 06-877-5111(ex.3115)

Email: kouasuke@imb.osaka-u.ac.jp
 Human Gene Signature, 3'-directed CDNA sequence. We are not
 submitting the same CDNA sequence redundantly to DBJ since 1993.
 For the abundance information of clones with this sequence in this
 library and as well as in other 3'-directed libraries, see
 http://www.imb.osaka-u.ac.jp/bodymap/. The sequences of the clones
 represented by this GS sequences is also found there.
 Location/Qualifiers
 1..48

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="Human adult (K.Okubo)"
 /dev_stage="adult"
 /note="Organ: blood; Vector: 1-9t-11; Site: 1; Eco-RI;
 Monocytes were prepared from blood by ficoll-hypaque,
 percoll and T cell rosetting purification steps (purity:
 96 %). mRNA was prepared from activated monocytes from a
 patient with rheumatoid arthritis. mRNA was reverse
 transcribed with MuLV. Using Eco-RI linkers CDNA was
 cloned into 1-9t-11 vector arms. The CDNA library was
 screened by differential hybridization using radioactively
 marked ss-CDNA from activated and non-activated
 monocytes."
 BASE COUNT 18 a 6 c 9 g 15 t
 ORIGIN

Alignment Scores:

Pred. NO.: 6.49e+03 Length: 48
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2,33% Indels: 0
 DB: 14 Gaps: 0

US-09-698-781-3 (1-258) x C21082 (1-48)

OY 123 SerGlnAlaIleGlnSer 128

DB 22 TCACAGCAATTCAGACT 5

RESULT 15

A1197659/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

A1197659 49 bp mRNA linear EST 14-OCT-1998
 uc46e09.r1 Soares-mammary-gland.NMIMG Mus musculus CDNA clone
 IMAGE:1494184.5' similar to TR:Q33563 Q33563 EATRO 164 KINETOPLAST
 ;, mRNA sequence.
 A1197659
 A1197659.1 GI:3750265
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 49)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Waterston, R., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:931788
 Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers

FEATURES
source

1. .49
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1494184"
/clone_1lb="Soares-mammary-gland_NMLMG"
/sex="female (lactating)"
/tissue-type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 26 a 13 c 10 g 0 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
6.64e+03	6.00	49	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.33%	Indels:	0
DB:	9	Gaps:	0

US-09-698-781-3 (1-258) x A1197659 (1-49)

OY 19 ValLeuNpheuVal 24
|||||

Db 42 GTGCTCCTGTTCTGCTG 25

RESULT 16

LOCUS A281329 50 bp mRNA linear EST 30-JUN-2000

DEFINITION 4A3A-P2A7-F Anopheles gambiae immune competent 4A3A Anopheles

ACCESSION A281329

VERSION A281329.1 GI:6929210

KEYWORDS EST

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.
1 (bases 1 to 50)
Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C., Donohue, M., Schultze, J., Benes, V., Bork, P., Ansoorge, W., Soares, M.B. and Kafatos, F.C.
Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines
Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)
Contact: Dimopoulos G
Fotis C. Kafatos Laboratory
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg, Germany.
Location/Qualifiers

FEATURES
source

1. .50
/organism="Anopheles gambiae"
/strain="4A r/r"
/db_xref="taxon:7165"
/clone="4A3A-P2A7"
/clone_1lb="Anopheles gambiae immune competent 4A3A"
/cell_line="Immune competent 4A3A"
/lab_host="E. coli DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; sequenced from

forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line Oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two approaches to Facilitate Gene Discovery, Genome Research 6, 791-806."

BASE COUNT 19 a 14 c 8 g 8 t 1 others

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
6.79e+03	6.00	50	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.33%	Indels:	0
DB:	9	Gaps:	0

US-09-698-781-3 (1-258) x A281329 (1-50)

OY 232 LeuTysLeuThrLeuThr 237
|||||

Db 1 CTAAATTAACCTCACT 18

RESULT 17

LOCUS AUI02979 50 bp mRNA linear EST 30-AUG-2001

DEFINITION AUI02979 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone

ACCESSION KAT10408, mRNA sequence.

VERSION AUI02979.1 GI:13552500

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 50)
Suzuki, Y., Talra, H., Tsunoda, T., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., Okubo, K., Sakaki, Y., Nakamura, Y., Suyama, A. and Sugano, S.
Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites
EMBO Rep. 2 (5), 388-393 (2001)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S.
Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
Location/Qualifiers

FEATURES
source

1. .50
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="KAT10408"
/clone_1lb="Sugano Homo sapiens cDNA library"
/note="Differential display comparison of untreated and dimethylsulfoxide treated 0937 cells"

BASE COUNT 1 a 24 c 14 g 11 t

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
6.79e+03	6.00	50	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.33%	Indels:	0
DB:	9	Gaps:	0

US-09-698-781-3 (1-258) x AUI02979 (1-50)

OY 116 SerSerAlaProSerSer 121

Fri Mar 14 14:00:07 2003

us-09-6988-781-3.oli.rst

Page 10

D6	33	TCCTCCGCCCGTGTCC	50
RESULT 18			
LOCUS	AA616601		
DEFINITION	AA616601 v66h08.r1 Barstead mouse proximal colon MFLRB6 Mus musculus cDNA clone IMAGE:1026207 5' similar to gb:U3751 Mouse (MOUSE), mRNA sequence.	51 bp mRNA linear EST 07-OCT-1997	
ACCESSION	AA616601		
VERSION	AA616601.1	GI:2503806	
KEYWORDS	EST. house mouse. Mus musculus		
SOURCE ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE AUTHORS	1 (bases 1 to 51) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Getzel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R. The Mashu-HMI Mouse EST Project Unpublished (1996)		
TITLE JOURNAL COMMENT	Contact: Marra M/Mouse EST project Mashu-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGJ:576983 Sege considered overall poor quality Trace primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 1.		
FEATURES	location/Qualifiers		
source	1..51 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone_image="IMAGE:1026207" /clone_lib="Barstead mouse proximal colon MFLRB6" /dev_stage="7 day juvenile" /lab_host="DH10B" /note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TTTAGCATCTGAGATGGAGCGGCCCGCCTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTGCATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library constructed by Bob Barstead."		
BASE COUNT	10 a 15 c 9 g 17 t		
ORIGIN			
Alignment Scores:	Pred. No.: 6.94e+03 Length: 51		
Score:	6.00 Matches: 6		
Percent Similarity:	100.00%	Conservative: 0	
Best Local Similarity:	100.00%	Mismatches: 0	
Query Match:	2.33%	Gaps: 0	
DB:	9		
US-09-698-781-3 (1-258) x AA616601 (1-51)			
OY	27	LoulouProserPhepro 32	
D6	25	TTCCTACCGCTATTCACA	42
RESULT 19			
BH218157/c			

```

LOCUS       BH218157                51 bp    DNA    linear    GSS 08-NOV-2001
DEFINITION  1006077005.2EL_X1 1006 - RescuemU Grid G Zea mays genomic, DNA
sequence.
ACCESSION   BH218157
KEYWORDS    BH218157.1  GI:16810590
SOURCE      GSS
ORGANISM    Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Palcoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 51)
AUTHORS     Walbot V.
TITLE        Maize genomic sequences found using engineered RescuemU transposon
JOURNAL      Unpublished (2001)
COMMENT     Contact: Walbot V
              Department of Biological Sciences
              Stanford University
              855 California Ave, Palo Alto, CA 94304, USA
              Tel: 650 723 2227
              Fax: 650 725 8221
              Email: walbot@stanford.edu
              Possible ligation site of ends cut by 2 different endonucleases.
              Reverse complemented post-ligation sequence from source sequence.
              Plate: 1006077 row: 27
              Class: transposon-tagged.
FEATURES
  source
    1..51
    /organism="Zea mays"
    /cultivar="mixed background W23/A18/873"
    /db_xref="taxon:4577"
    /clone_lib="1006 - RescuemU Grid G"
    /tissue_type="leaf"
    /dev_stage="adult"
    /lab_host="DH10B"
    /note="Organ: leaf; Vector: RescuemU (engineered from
    Bluescript backbone); Site.1: BamHI; Site.2: BglII;
    RescuemU is a 4.9 kb, modified maize Mu transposon
    designed to allow plasmid rescue from total genomic DNA.
    Mu elements insert preferentially into transcription
    units. For more information on RescuemU, go to the web
    site "www.zmdb.iastate.edu" and follow the links for
    RescuemU. Grid G was grown at Stanford in 2000. DNA was
    extracted from leaf punches, double digested using BamHI
    and BglII, and ligated to form circular plasmids. DH10B
    cells were transformed and then screened on LB plates with
    ampicillin."
BASE COUNT      9 a      14 c      20 g      8 t
ORIGIN
Alignment Scores:
  Pred. No.:      6.94e+03      Length:      51
  Score:          6.00          Matches:      6
  Percent Similarity: 100.00%    Conservative: 0
  Best Local Similarity: 100.00% Mismatches:    0
  Query Matchn:    2.33%        Indels:      0
  DB:              17          Gaps:         0
US-09-698-781-3 (1-258) x BH218157 (1-51)
QY      117 SerialProSeSerTripp 122
        |||||||ProSeSerTripp
        |||||||ProSeSerTripp
DB       37 TCTGCACCATCTCTCG 20
RESULT 20
LOCUS     B00049                52 bp    DNA    linear    GSS 13-JUL-1999
DEFINITION  B00049-100b6-u CSRL flow sorted Chromosome 11 specific cosmid Homo
sapiens genomic clone CSRL-100b6, DNA sequence.
ACCESSION   B00049
KEYWORDS    B00049.1  GI:1409327
SOURCE      GSS
ORGANISM    human.

```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 52)
 Evans, G.A., Burbee, D., Davies, C., Habner, L., Oliver, T., Gilbert, M.,
 Jones, D., Ward, T., Gillian, E., Schagemann, J., Probst, S., Harris,
 J., Deford, J., McFarland, J., Buzinski, K., Khan, M., Kupfer, K. and
 Garner, H.R.
 Genomic Sequence Sampled Map of Chromosome 11
 Unpublished (1996)
 Contact: Evans GA, Shane Probst
 Mcdermott Center for Human Growth and Development
 University of Texas Southwestern Medical Center At Dallas
 5323 Harry Hines Blvd, Dallas TX 75235-8591
 Tel: 214-648-1600
 Fax: 214-648-1666
 Email: gevas@utsw.swned.edu, shane@mcdermott.swned.edu
 Seq primer: 17
 Class: cosmid ends
 High quality sequence stop: 52.
 Location/Qualifiers
 1. 52
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSRU-100h6"
 /clone_1lb="CSRU flow sorted Chromosome 11 specific
 cosmid"
 /sex="female"
 /cell_type="chimeric hamster somatic cell hybrid"
 /note="Vector: SCOS-1; Human Chromosome 11 specific cosmid
 library prepared from flow sorted human Chromosome 11
 derived from Chinese Hamster Ovary (CHO) monochromosomal
 somatic cell hybrid, J1"
 BASE COUNT 10 a 10 g 25 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.09e+03 Length: 52
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 17 Gaps: 0
 US-09-698-781-3 (1-258) x B00049 (1-52)
 OY 26 gLyLeuLeuProSerPhe 31
 Db 23 GGCCTTGGCCGCTTTT 40
 RESULT 21
 DR16KIT/c 57 bp DNA linear GSS 06-JUN-2002
 LOCUS Danio rerio genomic clone Dkey-16k1, genomic survey sequence.
 DEFINITION AL733018
 VERSION AL733018.1 GI:21344182
 KEYWORDS GSS;
 zebrafish.
 SOURCE Danio rerio
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 57)
 Humphray, S.J., Huckle, E. and Hunt, S.E.
 Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
 humphraysanger.ac.uk Unpublished
 This sequence was generated from the T7 end of BAC 16k1. 16k1 is
 part of the Daniokey Pilot BAC Library created by R. Plasterk and
 N.V. Keygene.
 Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
 Location/Qualifiers

source 1. 57
 /organism="Danio rerio"
 /db_xref="taxon:7950"
 /clone="Dkey-16k1"
 /tissue_type="testis"
 /note="vector pindigBAC-536"
 BASE COUNT 11 a 17 c 12 g 17 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.85e+03 Length: 57
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 17 Gaps: 0
 US-09-698-781-3 (1-258) x DR16KIT (1-57)
 OY 24 ValAGLyLeuLeuPro 29
 Db 44 GTGGCTGGAGCTGCTGCCA 27
 RESULT 22
 AL758979/c 58 bp DNA linear GSS 18-JUN-2002
 LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-164E11-013275.
 DEFINITION genomic survey sequence.
 ACCESSION AL758979
 VERSION AL758979.1 GI:21497327
 KEYWORDS GSS;
 thale cress.
 SOURCE Arabidopsis thaliana
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsidae.
 1
 Strizhov, N., Li, Y., Rosso, M., Vlehoever, P., Dekker, K., Saedler, H.
 and Weisshaar, B.
 A pipeline for automated high-throughput generation of FSTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines
 Unpublished
 2
 Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
 A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics
 Unpublished
 3 (bases 1 to 58)
 Strizhov, N., Rosso, M., Li, Y. and Weisshaar, B.
 Direct Submission
 Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer
 Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion close to or within gene At4g30820. The
 sequences are generated at the MPI for Plant Breeding Research in
 the context of the GABI-Kat project. GABI-Kat is part of the German
 Plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.
 Location/Qualifiers
 1. 58
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="GK-164E11-013275"
 /clone_1lb="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana
 plants (T1) which were transformed with the T-DNA from
 vector pAC161. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant

similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed.

BASE COUNT 12 a 10 c 18 g 18 t

ORIGIN

Alignment Scores:

Pred. No.:	8e+03	Length:	58
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.33%	Indels:	0
DB:	17	Gaps:	0

US-09-698-781-3 (1-258) x AL758979 (1-58)

Oy 37 LysAspProAlaPheThr 42

Db 36 AAGGATCCGCTTTCACA 19

RESULT 23

AZ316082/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

Location/Qualifiers

1. 59

electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g14732114/gb1A129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 19 a 14 c 13 g 13 t

ORIGIN

Alignment Scores:

Pred. No.:	8.15e+03	Length:	59
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.33%	Indels:	0
DB:	17	Gaps:	0

US-09-698-781-3 (1-258) x AZ316082 (1-59)

Oy 204 CysAlaSerCysProAsp 209

Db 50 TGTCCTCTTCTTCGAT 33

RESULT 24

AZ313488/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

1. 60

Location/Qualifiers

electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g14732114/gb1A129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMDA2 (g114732114181a1f129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 23 a 0 c 37 g 0 t
ORIGIN

Alignment Scores:

Pred. No.:	8.3e+03	Length:	60
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.33%	Indels:	0
DB:	17	Gaps:	0

US-09-698-781-3 (1-258) x AZ13488 (1-60)

OY 27 LeuLeuprospherPro 32
|||||
DB 55 CTCCTCCCTCCTCCCT 38

RESULT 25

BC099525 62 bp mRNA linear EST 29-JAN-2001
LOCUS na353c09.x1 NCI-CGAP-Co27 Homo sapiens cDNA clone IMAGE:4205200 3',
DEFINITION mRNA sequence.

ACCESSION BC099525
VERSION BC099525.1 GI:12594842
KEYWORDS EST.

SOURCE

human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 62)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov

JOURNAL

Unpublished (1997)

COMMENT

CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.llnl.gov

Seq primer: -40up from GIBCO.
Location/Qualifiers

FEATURES

source

1..62
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4205200"
/clone_id="NCI-CGAP-Co27"
/tissue_type="adenocarcinoma (mucinous component)"
/lab_host="DH10B"
/note="Organ: colon; Vector: pAMP1; mRNA made from colonic
adenocarcinoma, cDNA made by oligo-dT priming.
Directionally cloned into udc sites. Size-selected on
agarose gel. Average insert size 300 bp. Primary library.
CDNA Library Preparation: David B. Kitzman, Ph.D.
Reference: Kitzman et al. (1996) Cancer Research
56:5360-5385."

BASE COUNT

17 a 16 c 12 g 17 t

ORIGIN

Alignment Scores:
Pred. No.: 8.61e+03
Score: 6.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 2.33%
DB: 12
Gaps: 0

US-09-698-781-3 (1-258) x BC099525 (1-62)

OY 252 AsnCysSerAsnSerIle 257
|||||
DB 10 AATGCTCGAATTCGATC 27

RESULT 26

BF037661 62 bp mRNA linear EST 20-OCT-2000
LOCUS 601462102P1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3865296 5',
DEFINITION mRNA sequence.

ACCESSION BF037661
VERSION BF037661.1 GI:10745972
KEYWORDS EST.

SOURCE

human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 62)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov

JOURNAL

Unpublished (1999)

COMMENT

Tissue Procurement: DCTP/DRP

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

Plate: LLM69608 row: d column: 01
High quality sequence start: 15
High quality sequence stop: 59.
Location/Qualifiers

FEATURES

source

1..62
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:3865296"
/clone_id="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site-1: NOTI;
Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT

14 a 18 c 19 g 11 t

ORIGIN

Alignment Scores:
Pred. No.: 8.61e+03
Score: 6.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 2.33%
DB: 12
Gaps: 0

US-09-698-781-3 (1-258) x BF037661 (1-62)

OY 19 VallLeuPhenylVal 24
|||||

DB 46 GTATTATGTTCTAGTA 29

RESULT 27

A2500262/c
 LOCUS A2500262 62 bp DNA linear GSS 05-OCT-2000
 DEFINITION IM0338H04R Mouse 10kb plasmid UUCG1M library Mus musculus genomic
 accession UUCG1M0338H04 R. DNA sequence.
 VERSION A2500262
 KEYWORDS A2500262.1 GI:10679897
 SOURCE GSS.
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 62)
 Dun, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 CONTACT: Robert B. Weiss
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112 USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 038 row: H column: 04
 Seq primer: CACACGACGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 62.
 Location/Qualifiers
 1..62
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCG1M0338H04"
 /clone_1lb="Mouse 10kb plasmid UUCG1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Ti-resistant, F-"
 /note="Vector: PMD42nv. Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (911473114191AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 27 a 1 c 32 g 2 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 8.61e+03 Length: 62
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 17 Gaps: 0
 US-09-698-781-3 (1-258) x A2500262 (1-62)

QY 27 LeuLeuproserPhePro 32
 Db 53 CTTCTTCCTTCCTTCCT 36
 RESULT 28
 AA761487
 LOCUS AA761487 63 bp mRNA linear EST 07-FEB-1998
 DEFINITION n222H09.S1 NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1288577 3',
 mRNA sequence.
 accession AA761487
 version AA761487.1 GI:2810417
 keywords EST.
 source human.
 organism Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 63)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 contact: Robert Strausberg, Ph.D.
 Email: cgap@nci.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNLT at:
 www.bio.linn.gov/dbp/image/image.html
 insert length: 892 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 61.
 Location/Qualifiers
 1..63
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /db_xref="IMAGE:1288577"
 /clone="IMAGE:1288577"
 /clone_1lb="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: pTR73D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I. Site 2: Eco RI. 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Gerald Marti (CBER). CDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTCACATCTGTGAGTGGGAGGCGCCCTCATTTTCTTTT-3'
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTR73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 17 a 14 c 2 g 30 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 8.76e+03 Length: 63
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.33% Indels: 0
 DB: 9 Gaps: 0
 US-09-698-781-3 (1-258) x AA761487 (1-63)
 QY 27 LeuLeuproserPhePro 32
 Db 27 TTATTACCTTCCTTCCTCA 44

Fri Mar 14 14:00:07 2003

us-09-698-781-3.011.rst

Page 15

RESULT 29
AF219092/c 63 bp DNA linear GSS 17-APR-2000
LOCUS AF219092 Human Homo sapiens genomic clone V(Uc9), DNA sequence.
DEFINITION AF219092
ACCESSION AF219092
VERSION AF219092.1 GI:7581538
KEYWORDS GSS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 63)
AUTHORS Hamshire, M., Cross, S., Daniels, M., Lennon, G. and Brook, J.D.
TITLE A transcript map of a 10-Mb region of chromosome 19: A source of
genes for human disorders, including candidates for genes involved
in asthma, heart defects, and eye disorders
JOURNAL Genomics 63 (3), 425-429 (2000)
MEDLINE 20171383
COMMENT Contact: Hamshire M
Institute of Genetics
University of Nottingham
Queen's Medical Center, Nottingham, NG7 2LT, United Kingdom
Class: exon-trapped.
Location/Qualifiers
FEATURES
source 1. 63
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human"
/note="Vector: pMOS Blue"
BASE COUNT 20 a 8 c 25 g 10 t
ORIGIN
Alignment Scores:
Pred. No.: 8.76e+03 Length: 63
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.33% Indels: 0
DB: 17 Gaps: 0
US-09-698-781-3 (1-258) x AF219092 (1-63)
Oy 116 SerSerAlaProSerSer 121
Db 46 TCCCTCGCCCATCATCT 29
RESULT 30
BI388670 64 bp mRNA linear EST 14-DEC-2001
LOCUS BI388670
DEFINITION EST-CD34NN-038 cDNA library from human CD34+ stem/progenitor cells
ACCESSION BI388670
VERSION BI388670.1 GI:17737253
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 64)
AUTHORS Zhou, G., Chen, J., Lee, S., Terry, C., Rowley, J.D. and Wang, S.M.
TITLE The pattern of gene expression in human hematopoietic CD34+
stem/progenitor cells
JOURNAL Unpublished (2001)
COMMENT Contact: Wang SM
Hem/Onc
University of Chicago Medical Center
5841 S. Maryland Ave., MC2115, Chicago, IL 60637, USA
Tel: 773-702-6788
Fax: 773-702-3002
Email: swang1@midway.uchicago.edu
This EST fragment was amplified from cDNA library of human CD34+
stem/progenitor cells with GLDI technique (Generation of Longer

cDNA fragments from SAGE tags for Gene Identification, Proc. Natl.
Acad. Sci. USA 97, 349, 2000), which starts from the 3' end till
the first CATG site of the targeted cDNA sequence.
Seq primer: M13 Forward.

FEATURES
source 1. 64
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="cDNA library from human CD34+ stem/progenitor
cells"
/tissue_type="Bone marrow"
/cell_type="CD34+ stem/progenitor cells"
/note="3-ESTs converted from the SAGE tag sequences using
GLDI method"
BASE COUNT 8 a 21 c 16 g 19 t
ORIGIN
Alignment Scores:
Pred. No.: 8.92e+03 Length: 64
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.33% Indels: 0
DB: 13 Gaps: 0
US-09-698-781-3 (1-258) x BI388670 (1-64)
Oy 26 GlyLeuLeuProSerPhe 31
Db 26 GGGCTCTCTCCACGCTT 43

Search completed: March 14, 2003, 07:14:15
Job time : 1730 secs